CLAIMS

What is claimed is:

1. A method for clustering data comprising:

(a) receiving a plurality of data points for clustering;

(b) receiving a size parameter for specifying the number of data points to be moved at one time;

(c) clustering the data points by using the size parameter to generate clustered results;

(d) determining whether the clustered results are satisfactory;

(e) when the clustered results are satisfactory, stop clustering;

(f) otherwise when the clustered results are not satisfactory, revise the size parameter, perform clustering based on the revised size parameter and the clustered results, and proceed to step (d).

2. The method of claim 1 wherein step (c) further comprises:

(c1) evaluating subsets of data points in each cluster for moving into every other cluster by using a predetermined metric; wherein the number of data points in the subset is specified by the size parameter.

3. The method of claim 2 wherein step (c1) further comprises:

(c1_1) determining a geometric center of the subset of data points being evaluated for a move;

(c1_2) using the geometric center of the subset of data points in the predetermined metric to generate a value.

4. The method of claim 3 wherein step (c1) further comprises:

(c1_3) determining whether the value is greater than zero;

(c1_4) when the value is greater than zero, moving the subset of data points from a Move_From cluster to a Move_To cluster;

(c1_5) when the value is not greater than zero, determining if there are more subsets to evaluate;

(c1_6) when there are more subsets to evaluate, proceeding to step (c1);

(c1_7) when there are no more subsets to evaluate, determining whether any point has moved;

(c1_8) when a point has moved, proceeding to step (c1); and

(c1_9) when no point has moved, stopping the processing.

5. The method of claim 4 wherein each data point has a membership with one cluster; wherein step (c1_4) further comprises:

simultaneously updating the membership of at least two data points from the membership of the Move_From cluster to the membership of the Move_To cluster.

6. The method of claim 4 wherein step (c1_4) further comprises:

updating the count of the Move_From cluster;

updating the center of the Move_From cluster;.

updating the count of the Move_To cluster; and

updating the center of the Move_To cluster.

7. The method of claim 1 wherein revising the size parameter of step (f) further comprises:

(f_1) decreasing the size parameter.

8. The method of claim 1 wherein step (d) further comprises:

(d_1) employing a predetermined metric for determining whether the clustered results are satisfactory; wherein the predetermined metric includes a geometric center of the subset of points that are being evaluated for move.

9. The method of claim 8 wherein the predetermined metric includes the following expression:

where U is the subset of data points being evaluated for the move, $|U|$ is the size of U that is specified by the size parameter, m_{oo} is the geometric center of U, M_I and m_j are the centers of the clusters and n_i and n_j are the counts of the clusters.

10. The method of claim 1 wherein the clustering method is utilized in one of a data mining application, customer segmentation application, document categorization application, scientific data analysis application, data compression application, vector quantization application, and image processing application.

11. A clustering system comprising:

(a) a source of data points to be clustered; and

(b) an aggregated clustering module for moving at least two data points at one time between a Move_From cluster and a Move_To cluster;

wherein the aggregated clustering module includes

a move determination unit for evaluating the move of subsets of data points from each cluster to every other cluster and determining when such a move should be performed; and

an aggregated move unit coupled to the move determination unit that updates a Move_From count, Move_From center, a Move_To count, and a Move_To center.

12. The clustering system of claim 11 wherein the aggregated clustering module further comprises:

(a) a first input for receiving the data points;

(b) a second input for receiving initial center points;

(c) a third input for receiving a number of points to move at one time;

(d) a parameter for storing the center point associated with each cluster

(e) a parameter for storing the count of data points associated with each cluster;

wherein the data points, center points and counts, are utilized by the move determination unit for move evaluation and determination and by the aggregated move unit for count update and center update.

13. The clustering system of claim 11 wherein the move determination unit further comprises:

a geometric center determination unit for determining the geometric center of a current subset of data points and providing the geometric center to the move determination unit for move evaluation and move determination.

14. The clustering system of claim 13 wherein the move determination unit further comprises:

a move evaluation mechanism for employing a predetermined metric for move evaluation; wherein the predetermined metric includes the geometric center of a current subset of data points.

15. The clustering system of claim 11 is configured for one of a data mining application, customer segmentation application, document categorization application, scientific data analysis application, data compression application, vector quantization application, and image processing application.